УДК: 004.93 https://doi.org/10.15407/jai2024.04.256

В. М. Синєглазов[1], О. А. Похиленко[2]
[1,2]Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Україна
пр. Берестейський, 37, м. Київ, 03056
[1]svm@nau.edu.ua
[2]pokhilenko.alex@gmail.com
[1]https://orcid.org/0000-0002-3297-9060
[2]https://orcid.org/0000-0002-1562-2051

ІНТЕЛЕКТУАЛЬНА МЕДИЧНА СИСТЕМА ДІАГНОСТИКИ ДЕФОРМАЦІЇ МІЖХРЕБЦЕВИХ ДИСКІВ

V. Sineglazov[1], O. Pokhylenko[2]
[1,2]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine
37, Beresteiskyi ave, Kyiv, 03056
[1]svm@nau.edu.ua
[2]pokhilenko.alex@gmail.com
[1]https://orcid.org/0000-0002-3297-9060
[2]https://orcid.org/0000-0002-1562-2051

INTELLIGENT MEDICAL SYSTEM FOR DIAGNOSIS OF INTERVERTEBRAL DISC DEFORMATION

Анотація. У статті розглядається застосування методів напівкерованого навчання та комп'ютерного зору для сегментації МРТ-зображень хребта та діагностики деформації міжхребцевих дисків. Оглянуто існуючі архітектури нейронних мереж для сегментації МРТ-зображень хребта та методи напівкерованого навчання, які використовуються при сегментації медичних зображень. Запропоновано систему діагностики деформації міжхребцевих дисків, яка складається з двох модулів: модуля сегментації та модуля діагностування. Наведено реалізацію запропонованої системи, що використовує дві згорткові нейронні мережі: U-Net для сегментації МРТ-зображень хребта та ResNet для класифікації ступеня деформації кожного міжхребцевого диска на основі класифікації дегенеративних змін за Pfirrmann. Розроблено програмну реалізацію медичної системи діагностики на мові програмування Python з використанням бібліотеки PyTorch. Виконано навчання нейронних мереж на відкритому наборі даних МРТ-зображень хребта з використанням різновиду методу напівкерованого навчання Mean Teacher. У результаті перевірки було виявлено, що система здатна виконувати сегментацію з високою точністю. Виявлено, що точне визначення класу дегенеративних змін за Pfirrmann залишається складним завданням, однак введення іншої класифікації дозволило підвищити точність діагностики деформації міжхребцевих дисків. Запропонована медична система передбачає додавання нових модулів діагностування, що робить можливим її застосування для комплексного аналізу різних захворювань хребта.

Ключові слова: згорткова нейронна мережа, МРТ хребта, деформація міжхребцевих дисків, напівкероване навчання, комп'ютерний зір, сегментація зображень, класифікація зображень.

Abstract. The article examines the application of semi-supervised learning and computer vision methods for segmentation of spine MR images and diagnosis of deformation of intervertebral discs. Existing neural network architectures for spine MR image segmentation and semi-supervised learning methods used in medical image segmentation are reviewed. A system for diagnosis deformation of intervertebral discs is proposed, which consists of two modules: a segmentation module and a diagnostic module. The implementation of the proposed system using two convolutional neural networks is presented: U-Net for segmentation of spine MR images and ResNet for classification of the degree of deformation of each intervertebral disc based on the Pfirrmann classification of degenerative changes. The software implementation of the medical diagnosis system was developed in the Python programming language using the PyTorch library. Neural networks were trained on an open dataset of spine MR images using a variant of the Mean Teacher semi-supervised learning method. As a result of the verification, it was found that the system is capable of performing segmentation with high accuracy. It was found that the exact prediction of the grade of degenerative changes according to Pfirrmann remains a difficult task, but the introduction of another classification made it possible to increase the accuracy of diagnosis of deformation of intervertebral discs. The proposed medical system involves the addition of new diagnostic modules, which makes it possible to use it for the comprehensive analysis of various spine diseases.

Keywords: convolutional neural network, spine MR images, deformation of intervertebral discs, semi-supervised learning, computer vision, image segmentation, image classification.

Вступ

Сегментація МРТ-зображень, зокрема МРТ-зображень хребта, займає ключове місце в сучасній медичній діагностиці та лікуванні різних захворювань. Основною задачею сегментації є позначення подібних областей (сегментів) у зображенні на рівні пікселів. У випадку сегментації зображень хребта важливими елементами є тіла хребців, міжхребцеві диски та хребетний канал. Коректна ідентифікація та визначення положення і форми цих структур мають важливе значення при діагностиці хребетних патологій та плануванні лікувальних заходів.

В діагностиці дегенеративних змін хребта часто використовується класифікація за Pfirrmann, яка дозволяє оцінити ступінь дегенерації міжхребцевих дисків. Для точної автоматичної діагностики деформації необхідна якісна сегментація, оскільки правильне визначення контурів і форм міжхребцевих дисків впливає на точність оцінки їх стану. Сегментація також може бути використана при виявленні таких патологій, як спондилолістез, грижа, випинання та звуження дисків тощо.

Традиційна ручна сегментація зображень є трудомістким процесом, що вимагає значних зусиль і фахових навичок. Проте останніми роками спостерігається зростання популярності автоматизованих методів сегментації, зокрема, з використанням згорткових нейронних мереж.

Автоматична сегментація медичних зображень стикається з рядом складнощів. Медичні зображення, зокрема МРТ, часто мають низьку контрастність та підвищений рівень шуму, що ускладнює точне виділення границь об’єктів. Крім цього, подібність між різними класами об’єктів на зображенні може перешкоджати коректній класифікації. Проблема дисбалансу даних також є актуальною для медичних зображень, оскільки деякі класи можуть бути недостатньо представлені у вибірках. Також слід зазначити, що в реальних умовах кількість немаркованих зображень може значно перевищувати кількість маркованих, що робить використання методів напівкерованого навчання дуже бажаним. Всі ці проблеми роблять розробку високоточної медичної системи діагностики нетривіальною задачею.

Постановка проблеми

У рамках дослідження основним завданням була розробка медичної системи діагностики деформації міжхребцевих дисків. Це включало навчання моделей штучних нейронних мереж для автоматичної сегментації МРТ-зображень хребта та класифікації деформацій окремих дисків. При цьому навчання виконувалося на відкритому наборі даних з використанням методів напівкерованого навчання.

Як вже зазначалося, через низький контраст і високий рівень шуму в МРТ-зображеннях, автоматичне виявлення контурів дисків і хребців може бути складним завданням. Це є особливо важливим, оскільки точність сегментації має безпосередній вплив на якість наступної класифікації дегенеративних змін.

Тому важливо використовувати відповідні моделі та методи, які можуть дозволити точно сегментувати та розпізнавати структури, навіть у складних умовах. Для цього були поставлені наступні завдання:

1. Визначення, які архітектури нейронних мереж застосовуються для сегментації МРТ-зображень.
2. Вибір методу напівкерованого навчання серед низки методів, які використовуються при сегментації медичних зображень.
3. Навчання нейронних мереж.
4. Проведення експериментальних досліджень та аналіз результатів.

Аналіз останніх досліджень і публікацій

Відповідно до [1], за 2015-2021 роки використання глибокого навчання в аналізі зображень хребта розглядалося в 79 публікаціях, причому в останні роки спостерігалося зростання кількості

досліджень у цій галузі. МРТ-зображення є найбільш популярними – вони використовувалися в 40.5% публікацій. Задачі діагностування та сегментації є найпоширенішими: задача сегментації вирішувалася в 30.4% від усіх публікацій, а задача діагностування – в 35.4%. Найбільш популярною архітектурою для сегментації медичних зображень виявилася U-Net та її модифікації.

В результаті пошуку літератури були знайдені публікації [2-6], в яких розглядалися задачі сегментації та діагностування по МРТ-зображенням. В [2] використовувалися архітектури U-Net++ та Yolov5x, також зазначалося, що через нестачу тренувальних даних була помічена схильність нейронних мереж до перенавчання навіть у випадку застосування аугментації даних. В [3] пропонується використання послідовного умовного навчання з підкріпленням (Sequential Conditional Reinforcement Learning), та трьох мереж: мережі анатомічного моделювання для фокусування уваги, повнозв'язної залишкової мережі ResNet для виявлення обмежувальної рамки, та мережі Y-Net для сегментації. В [4] розглядається Spine Explorer – програма, що використовує U-Net для сегментації тіл хребців, міжхребцевих дисків та хребетного каналу. Ця програма цікава тим, що дозволяє не лише виконувати сегментацію МРТ-зображень хребта, але й виконувати обчислення певних показників, які можуть бути корисними в діагностиці. В [5] пропонується MMCL-Net – багатозадачна багатоструктурна мережа кореляційного навчання, що дозволяє одночасно виявляти, сегментувати та класифікувати (визначати ступінь пошкодження) тіла хребців, міжхребцеві диски та хребетний канал. Ця мережа дозволяє одночасно з сегментацією виконувати класифікацію частин хребта за ступенем ураження (нормальний стан або легкі, виражені чи важкі ураження). В [6] розглядається система для мультимодальної двовимірної сегментації тіл хребців і міжхребцевих дисків, що може виконувати сегментацію МРТ, КТ та рентгенівських знімків, в основі системи лежить адаптована архітектура Mask R-CNN.

Окрім публікацій, присвячених задачам сегментації та діагностування, також були розглянуті роботи [7-9], які присвячені використанню методів напівкерованого навчання щодо сегментації медичних зображень. В [7] пропонується новий метод – кореляційне взаємонавчання для сегментації медичних зображень, яке засновано на перехресному псевдоконтролі. В [8] пропонується ще один метод напівкерованого навчання для сегментації медичних зображень – MLB-Seg та алгоритм покращення псевдоміток. В [9] пропонується модифікація методу Mean Teacher за рахунок використання двонаправленого копіювання маркованих та немаркованих даних. Зазначається, що використання цього підходу дозволяє в певній мірі запобігти перенавчанню на малих вибірках, що є актуальним для відкритих наборів даних медичних зображень, які зазвичай є не дуже великими.

Мета дослідження

Метою даного дослідження є розробка автоматизованої медичної системи для діагностики деформації міжхребцевих дисків на основі сегментації МРТ-зображень. Враховуючи розглянуті публікації, для досягнення цієї мети було вирішено розробити систему у вигляді двох модулів: модуля сегментації та модуля діагностування. Таким чином, необхідно було виконати наступні завдання:

1. Реалізувати модуль сегментування МРТ-зображень хребта на основі нейронної мережі U-Net [10].
2. Застосувати до U-Net варіацію методу напівкерованого навчання Mean Teacher [11], враховуючи проблеми з перенавчанням у випадках обмежених тренувальних даних.
3. Реалізувати модуль діагностування для визначення ступенів деформації окремих міжхребцевих дисків на основі нейронної мережі ResNet.
4. Провести експериментальне дослідження для оцінки ефективності

розроблених моделей на відкритому наборі даних [12].

Таким чином, виконання цих поставлених завдань буде свідчити про можливість створення та потенціальне застосування інтелектуальної медичної системи для діагностики деформації міжхребцевих дисків на основі сегментації МРТ-зображень.

Виклад основного матеріалу

U-Net є однією з найбільш відомих архітектур для сегментації медичних зображень [10]. Ця архітектура складається з кодувальника і декодувальника. Кодувальник виділяє ознаки шляхом застосування шарів згортки та Max Pooling, зменшуючи роздільну здатність, а декодувальник збільшує роздільну здатність виділених ознак. Також в U-Net застосовуються пропускні з'єднання на відповідних рівнях роздільної здатності, що дозволяє зберігати просторову інформацію. В якості функції активації пропонується використовувати ReLU. Варто зазначити, що існують різновиди мережі U-Net, в яких виконується нормалізація пакетів після кожного згорткового шару. Саме такий варіант архітектури було вирішено використати в цьому дослідженні.

В класичному методі напів-керованого навчання Mean Teacher використовуються дві копії однієї нейронної мережі: вчитель та учень [11]. Мережа-вчитель генерує псевдомітки для немаркованих даних, а мережа-учень навчається на маркованих та псевдомаркованих даних. На кожному кроці виконується навчання мережі-учня на аугментованих даних. Після того, як було виконано оновлення вагових коефіцієнтів мережі-учня, вагові коефіцієнти мережі-вчителя оновлюються як експоненціальне ковзне середнє вагових коефіцієнтів учня. В цьому дослідженні використовувалася варіація цього методу з використанням двонаправленого копіювання та вставки [9]. Ця варіація полягає в генерації вхідних даних для мережі-студента шляхом змішування маркованих та псевдомаркованих частин зображень. Відповідно до [9], для підвищення точності моделі замість використання мереж з випадковою ініціалізацією ваг, рекомендується виконати попереднє навчання мережі-вчителя на маркованому наборі даних. При цьому дані повинні бути аугментовані шляхом копіювання та вставки сегментів.

Під час навчання та перевірки моделей використовувався набір даних [12], який містить 447 сегментованих МРТ-зображень хребта, при цьому 360 зображень використовувалися для тренування і 87 зображень – для тестування. Сегментація містила тіла хребців, міжхребцеві диски та хребетний канал. Цей набір даних також містив радіологічні оцінки, включаючи наявність дегенеративних змін за класифікацією Pfirrmann, розподіл яких можна побачити в таблиці 1.

Таблиця 1. Розподіл радіологічних оцінок

Pfirr-mann	Тренувальна вибірка		Тестувальна вибірка	
	Диски (1240)	Пацієнти (179)	Диски (280)	Пацієнти (39)
I	241 (19.4%)	66 (36.9%)	45 (16.1%)	16 (41.0%)
II	260 (21.0%)	106 (59.2%)	81 (28.9%)	14 (35.9%)
III	348 (28.1%)	133 (74.3%)	70 (25.0%)	30 (76.9%)
IV	240 (19.4%)	92 (51.4%)	51 (18.2%)	25 (64.1%)
V	151 (12.2%)	79 (44.1%)	51 (18.2%)	16 (41.0%)

Приклад 2D зрізу МРТ-зображення разом із масками для сегментації (тіла хребців, хребетний канал та міжхребцеві диски) можна побачити на рис. 1.

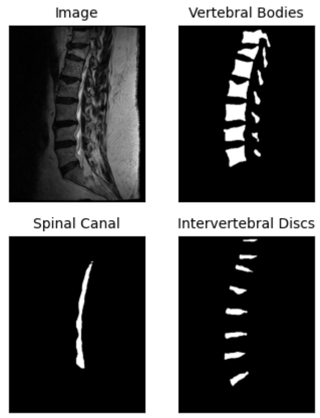


Рис. 1. Приклад зображення та масок з [12].

При розробці програмного забезпечення використовувалася мова програмування Python, а також бібліотеки PyTorch, NumPy, Pandas, scikit-learn, OpenCV та інші. Розроблена програма дозволяє навчити декілька моделей.

Спочатку розглянемо моделі, які стосуються модуля сегментації. Створюється одна мережа для сегментації і виконується її кероване навчання – ця модель використовується лише для порівняння результатів з відповідними результатами при застосуванні напівкерованого навчання. Далі готується новий тренувальний набір даних, в якому більшість даних є немаркованими (90%). Цей набір використовується для моделювання умов, коли доступні лише обмежені марковані дані, що є типовим сценарієм у медичних дослідженнях. Це дозволяє перевірити ефективність підходу в реальних умовах, де велика кількість зображень не має відповідних анотацій. Таким чином, частково маркований набір застосовується для напівкерованого навчання. Далі створюється мережа-вчитель, виконується її попереднє навчання на маркованій частині набору тренувальних даних. Далі створюється мережа-учень з ваговими коефіцієнтами, які є точною копією вагових коефіцієнтів мережі-вчителя. Після цього застосовується метод Mean Teacher з двонаправленим копіюванням та вставкою. В результаті цього отримується мережа, навчена за допомогою напівкерованого навчання, що буде демонструвати результати, які повинні бути кращими за початкові результати мережі-вчителя після попереднього навчання, але зазвичай ці результати будуть гіршими, ніж результати від повністю керованої моделі.

Навчання кожної моделі для сегментації виконувалося на 40 епохах з використанням оптимізатору Adam, початкова швидкість навчання: 0.001 зі зменшенням в 10 разів кожні 20 епох, розмір пакета дорівнював 8. В якості функції втрат було використано JaccardLoss – ця функція заснована на індексі Жаккара (перетині над об'єднанням). Після кожної епохи зберігалася модель з найвищим значенням коефіцієнта подібності Дайса (F_1-міри) на тестувальній вибірці. Метрики моделей для сегментації з кожного етапу навчання можна побачити в таблиці 2, де етап «кероване навчання» відповідає моделі, яка навчалася на повністю маркованому наборі даних; «попереднє навчання» – моделі, яка навчалася на 10% маркованих даних, а «напівкероване навчання» – моделі, яка навчалася зі застосуванням методу Mean Teacher з двонаправленим копіюванням та вставкою та 90% немаркованими даними.

Таблиця 2. Метрики моделі для сегментації

Метрики / Етап	Індекс Жаккара	Індекс Дайса
Кероване навчання	0.835	0.911
Попереднє навчання	0.196	0.711
Напівкероване навчання	0.799	0.888

З цієї таблиці можна побачити, що за допомогою напівкерованого навчання обраним методом при 10% маркованих даних в наборі вдалося отримати значно кращі результати, ніж при навчанні керованим підходом на цих 10% даних. Як видно, отримані результати достатньо близькі до результатів, які були отримані при навчанні зі 100% маркованим набором даних: індекс Жаккара відрізняється лише на 0.036, а коефіцієнт подібності Дайса – на 0.023. Отже, метод напівкерованого навчання Mean Teacher з двонаправленим копіюванням та вставкою продемонстрував чудові результати. Результат сегментації цією моделлю (нижній рядок) в порівнянні з еталонними масками (верхній рядок) можна побачити на рис. 2.

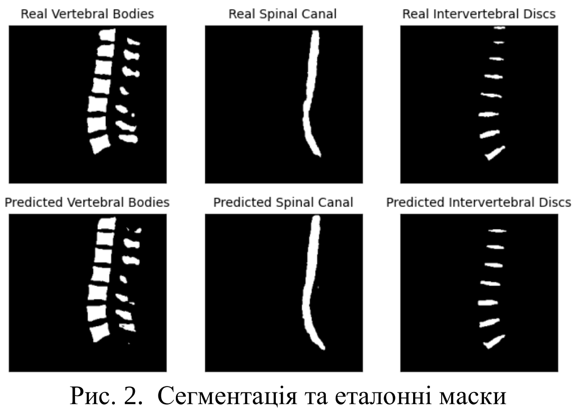


Рис. 2. Сегментація та еталонні маски

Тепер розглянемо модуль діагностування. Він застосовується після того, як відпрацює модуль сегментації, оскільки на вхід нейронної мережі для діагностування подається не тільки початкове зображення, але й результат сегментації. Через те, що в медичній системі діагностики виконується визначення деформації кожного міжхребцевого диска окремо, на вхідних даних цього модуля за допомогою бібліотеки OpenCV виконується виділення контурів сегментованих дисків, після цього відповідні контурам регіони вирізаються, що дозволяє отримати набір невеликих регіонів інтересу, кожен з яких містить один міжхребцевий диск. Далі ці регіони подаються на вхід нейронної мережі, що виконує класифікацію деформації диска. Приклад вхідних даних можна побачити на рис. 3 – на вхід подаються 3 канали: початкове МРТ-зображення, а також 2 канали від модуля сегментації (тіла хребців та міжхребцеві диски).

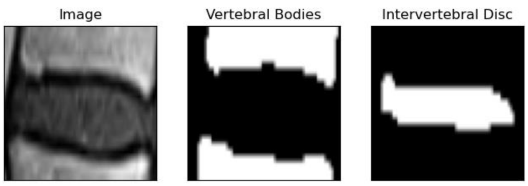


Рис. 3. Вхідні дані модуля діагностування

Для задачі класифікації було вирішено використовувати архітектуру ResNet-18. Навчання моделей для класифікації виконувалося на 45 епохах з використанням оптимізатора Adam, початкова швидкість навчання: 0.001 зі зменшенням в 10 разів кожні 15 епох, розмір пакета дорівнював 32. В якості функції втрат було використано FocalLoss зі значенням γ=2 – ця функція втрат спеціально розроблена для покращення навчання в умовах дисбалансу класів, зосереджуючи увагу на важких зразках, що дозволяє моделі краще справлятися з класами, представленими в обмеженій кількості. Після кожної епохи зберігалася модель з найвищим значенням точності на тестувальній вибірці, після чого для найкращої моделі будувалася матриця невідповідностей.

Перший експеримент був спрямований на передбачення ступеня дегенеративних змін міжхребцевих дисків за класифікацією Pfirrmann, що включає здоровий стан та п’ять рівнів дегенеративних змін. На рис. 4 можна побачити графік залежності точності моделі (accuracy) від епохи.

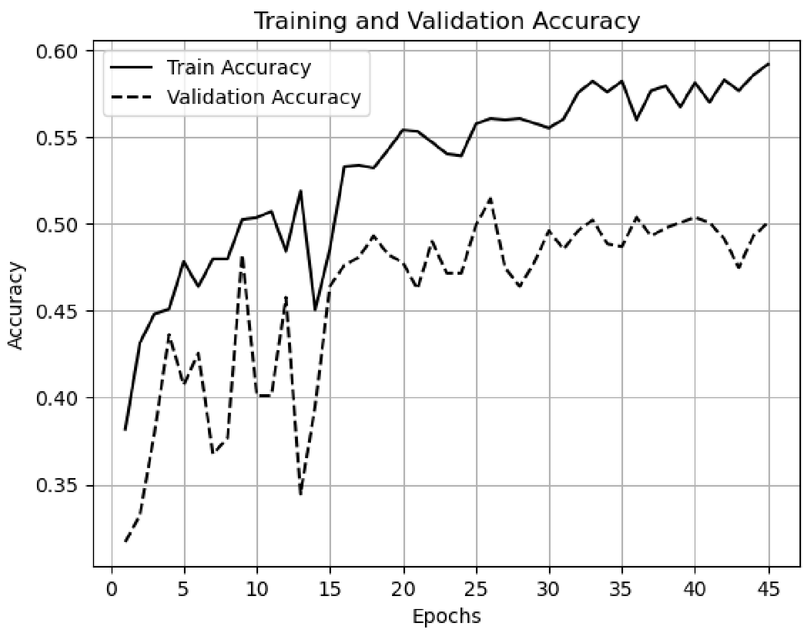


Рис. 4. Експеримент №1, точність моделі

На останніх епохах спостерігається стабілізація точності на тестовій вибірці (≈0.5), в той же час точність на тренувальній вибірці продовжує рости, що може свідчити про перенавчання моделі. На тестовій вибірці значення F_1-міри дорівнювало ≈0.50, однак модель значно краще справлялася з класифікацією V та IV рівнів за Pfirrmann: повнота (recall) для цих рівнів дорівнює 0.74 та 0.61 відповідно. Матриця невідповідностей найкращої моделі зображена на рис. 5. На цій матриці можна побачити цікаву особливість: модель схильна переоцінювати рівні за Pfirrmann, а також має певні проблеми із суміжними класами (наприклад, рівні III-IV та IV-V), що пояснюється подібністю між класами. Також можна помітити, що в обраному наборі даних були майже відсутні здорові міжхребцеві диски.

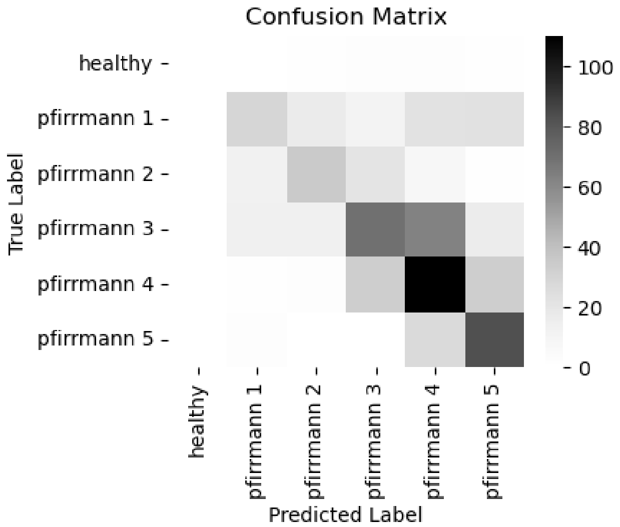


Рис. 5. Експеримент №1, Confusion Matrix

У другому експерименті реалізовано власну класифікацію, що поділяє диски на три категорії: незначна деформація (рівні за Pfirrmann до II включно), помірна деформація (рівень III) та тяжка деформація (рівні IV–V). Відповідний графік залежності точності моделі (accuracy) від епохи зображено на рис. 6. Як і під час минулого експерименту, на останніх епохах вже спостерігається перенавчання, точність на тестовій вибірці коливалася біля значення ≈0.65, після чого почала падати.

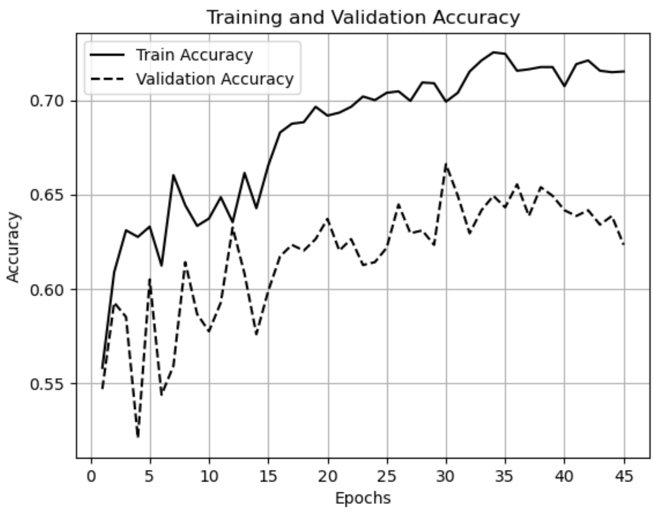


Рис. 6. Експеримент №2, точність моделі

На рис. 7 зображена матриця невідповідностей для цього експерименту. Помітно, що модель значно краще виявляє важкі та незначні деформації міжхребцевих дисків, ніж помірні.

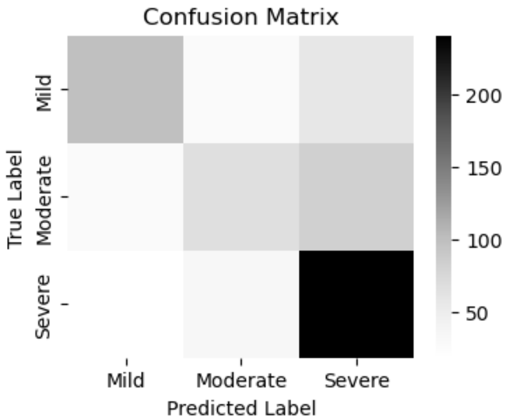


Рис. 7. Експеримент №2, Confusion Matrix

В таблиці 3 наведені метрики для кожного з класів окремо.

Запропонована класифікація деформації міжхребцевих дисків є лише прикладом використання модуля діагностування розробленої медичної системи. Варто зазначити, що ця система може бути адаптована до інших клінічних сценаріїв, таких як виявлення інших типів патологій хребта, за рахунок розробки додаткових модулів діагностування.

Таблиця 3. Метрики моделі діагностування

Метрики / Класи	Влучність (precision)	Повнота (recall)	F_1-міра
Незначна	0.69	0.54	0.60
Помірна	0.53	0.38	0.44
Тяжка	0.63	0.83	0.72

Висновки

В даному дослідженні запропоновано модульну інтелектуальну медичну систему діагностики деформації міжхребцевих дисків по МРТ-зображеннях хребта. Система складається з двох модулів: модуля сегментації та модуля діагностування. Було реалізовано два модуля системи, що використовують згорткові нейронні мережі: U-Net для сегментації МРТ-зображень хребта та ResNet для класифікації ступеня деформації кожного міжхребцевого диска відповідно. Нейронні мережі були навчені на відкритому наборі даних МРТ-зображень хребта, при цьому було застосовано варіацію методу напівкерованого навчання Mean Teacher з двонаправленим копіюванням та вставкою. У результаті перевірки було виявлено, що система здатна виконувати сегментацію з відносно високою точністю навіть при навчанні за допомогою напівкерованого навчання при 10% маркованих даних в наборі. Під час дослідження виявлено, що точне визначення класу дегенеративних змін за Pfirrmann є складним завданням, однак введення іншої класифікації (незначна, помірна та тяжка деформація дисків) дозволило підвищити точність діагностики деформації міжхребцевих дисків. Запропонована медична система має потенціал для подальшого розвитку завдяки можливості розширення її функціональності без потреби в перенавчанні мережі для сегментації. Це може досягатися за рахунок додавання нових модулів діагностування інших захворювань хребта.

Література

1. Qu, B., Cao, J., Qian, C., Wu, J., Lin, J., Wang, L., & Qu, X. (2022). Current development and prospects of deep learning in spine image analysis: a literature review. *Quantitative Imaging in Medicine and Surgery*, 12(6), 3454.

2. Guinebert, S., Petit, E., Bousson, V., Bodard, S., Amoretti, N., & Kastler, B. (2022). Automatic semantic segmentation and detection of vertebras and intervertebral discs by neural networks. *Computer Methods and Programs in Biomedicine Update*, 2, 100055.

3. Zhang, D., Chen, B., & Li, S. (2021). Sequential conditional reinforcement learning for simultaneous vertebral body detection and segmentation with modeling the spine anatomy. *Medical image analysis*, 67, 101861.

4. Huang, J., Shen, H., Wu, J., Hu, X., Zhu, Z., Lv, X., ... & Wang, Y. (2020). Spine Explorer: a deep learning based fully automated program for efficient and reliable quantifications of the vertebrae and discs on sagittal lumbar spine MR images. *The Spine Journal*, 20(4), 590-599.

5. Hong, Y., Wei, B., Han, Z., Li, X., Zheng, Y., & Li, S. (2020). MMCL-Net: spinal disease diagnosis in global mode using progressive multi-task joint learning. *Neurocomputing*, 399, 307-316.

6. Suri, A., Jones, B. C., Ng, G., Anabaraonye, N., Beyrer, P., Domi, A., ... & Rajapakse, C. S. (2021). A deep learning system for automated, multi-modality 2D segmentation of vertebral bodies and intervertebral discs. *Bone*, 149, 115972.

7. Gao, S., Zhang, Z., Ma, J., Li, Z., & Zhang, S. (2023, October). Correlation-Aware Mutual Learning for Semi-supervised Medical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 98-108.

8. Wei, Q., Yu, L., Li, X., Shao, W., Xie, C., Xing, L., & Zhou, Y. (2023, October). Consistency-Guided Meta-learning for Bootstrapping Semi-supervised Medical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 183-193.

9. Bai, Y., Chen, D., Li, Q., Shen, W., & Wang, Y. (2023). Bidirectional copy-paste for semi-supervised medical image segmentation. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11514-11524.

10. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference*, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.

11. Tarvainen, A., & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30.

12. Van der Graaf, J. W., van Hooff, M. L., Buckens, C. F., Rutten, M., van Susante, J. L., Kroeze, R. J., ... & Lessmann, N. (2024). Lumbar spine segmentation in MR images: a dataset and a public benchmark. *Scientific Data*, 11(1), 264.

References

1. Qu, B., Cao, J., Qian, C., Wu, J., Lin, J., Wang, L., & Qu, X. (2022). Current development and prospects of deep learning in spine image analysis: a literature review. *Quantitative Imaging in Medicine and Surgery*, 12(6), 3454.

2. Guinebert, S., Petit, E., Bousson, V., Bodard, S., Amoretti, N., & Kastler, B. (2022). Automatic semantic segmentation and detection of vertebras and intervertebral discs by neural networks. *Computer Methods and Programs in Biomedicine Update*, 2, 100055.

3. Zhang, D., Chen, B., & Li, S. (2021). Sequential conditional reinforcement learning for simultaneous vertebral body detection and segmentation with modeling the spine anatomy. *Medical image analysis*, 67, 101861.

4. Huang, J., Shen, H., Wu, J., Hu, X., Zhu, Z., Lv, X., ... & Wang, Y. (2020). Spine Explorer: a deep learning based fully automated program for efficient and reliable quantifications of the vertebrae and discs on sagittal lumbar spine MR images. *The Spine Journal*, 20(4), 590-599.

5. Hong, Y., Wei, B., Han, Z., Li, X., Zheng, Y., & Li, S. (2020). MMCL-Net: spinal disease diagnosis in global mode using progressive multi-task joint learning. *Neurocomputing*, 399, 307-316.

6. Suri, A., Jones, B. C., Ng, G., Anabaraonye, N., Beyrer, P., Domi, A., ... & Rajapakse, C. S. (2021). A deep learning system for automated, multi-modality 2D segmentation of vertebral bodies and intervertebral discs. *Bone*, 149, 115972.

7. Gao, S., Zhang, Z., Ma, J., Li, Z., & Zhang, S. (2023, October). Correlation-Aware Mutual Learning for Semi-supervised Medical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 98-108.

8. Wei, Q., Yu, L., Li, X., Shao, W., Xie, C., Xing, L., & Zhou, Y. (2023, October). Consistency-Guided Meta-learning for Bootstrapping Semi-supervised Medical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 183-193.

9. Bai, Y., Chen, D., Li, Q., Shen, W., & Wang, Y. (2023). Bidirectional copy-paste for semi-supervised medical image segmentation. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11514-11524.

10. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference*, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.

11. Tarvainen, A., & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30.

12. Van der Graaf, J. W., van Hooff, M. L., Buckens, C. F., Rutten, M., van Susante, J. L., Kroeze, R. J., ... & Lessmann, N. (2024). Lumbar spine segmentation in MR images: a dataset and a public benchmark. *Scientific Data*, 11(1), 264.

The article has been sent to the editors 30.09.24.
After processing 10.10.24.
Submitted for printing 30.12.24.